James Thornton
Internet Business Consultant

Browse the web faster. Get Firefox with Google Toolbar.
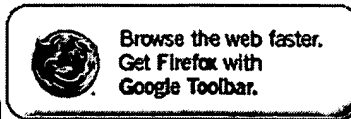
Google™

○ Web ◉ jamesthornton.c

| About James | Services | Writings | Documentation | Contact Me |

sort by date added, data published, title | **add paper**

Collaborative Filtering Research Papers

Summary: 52 abstracts with links to the full papers and reader comments.

A Bayesian Model for Collaborative Filtering (**2000**), *by Yung-Hsin Chen and Edward I. George (Department of MSIS, University of Texas at Austin)*

Consider the general setup where a set of items have been partially rated by a set of judges, in the sense that not every item has been rated by every judge. For this setup, we propose a Bayesian approach for the problem of predicting the missing ratings from the ob-served ratings. This approach incorporates similarity by assuming the set of judges can be partitioned into groups which share the same ratings probability distribution. This leads to a predictive distribution of missing ratings based on the posterior distribution of the groupings and associated ratings proba-bilities. Markov chain Monte Carlo methods and a hybrid search algorithm are then used to obtain predictions of the missing ratings.
bevo2.bus.utexas.edu/GeorgeE/Research%20papers/Bcollab.pdf - reader comments
Added by James Thornton on 2001-01-31

A Collaborative Filtering Agent System for Dynamic Virtual Communities on the Web (**1998**), *by O. de Vel and S. Nesbitt (Department of Computer Science, James Cook University)*

Collaborative filtering automatically retrieves and filters documents by considering the recommendations or feedback given by other users to the documents. In this paper we describe the webCobra recommendation system for automatically recommending high-quality web documents to users with similar interests on arbitrarily narrow information domains. User-centric virtual communities consisting of members whose recommendations have been deemed to be highly relevant with respect to a particular information domain will be automatically formed. We present some preliminary results and show that virtual collaborative communities defined by webCobra are able to dynamically modify their boundaries to allow for changes in user interests.
citeseer.nj.nec.com/de-collaborative.html - reader comments
Added by James Thornton on 2001-01-31

A Java-Based Approach to Active Collaborative Filtering (**1998**), *by Christopher Lueg and Christoph Landolt (AI-Lab, Department of Computer Science, University of Zurich)*

In this paper, we present a collaborative filtering approach to webpage filtering. The system supports users in exchanging recommendations and exploits the social relation between recommenders and recipients of recommendations instead of computing a degree of interest. In order to help users estimate the potential interestingness of a recommended webpage, the system augments the recommendation object with additional data indicating how previous recipients of the recommendation have dealt with the corresponding webpage. The system has been implemented as a collection of personal user agents exchanging recommendations with a central recommendation server. The user agents are implemented as Java applets and the recommendation server is a Java remote object realized as object factory.
www.ifi.unizh.ch/~lueg/abstracts/chi98late.html - reader comments
Added by James Thornton on 2001-01-31

Analysis of Recommendation Algorithms for E-Commerce (**2000**), *by Badrul Sarwar, George Karypis,*

Joseph Konstan, and John Riedl (GroupLens Research Group / Army HPC Research Center, Department of Computer Science and Engineering University of Minnesota)

Recommender systems apply statistical and knowledge dis- covery techniques to the problem of making product recom- mendations during a live customer interaction and they are achieving widespread success in E-Commerce nowadays. In this paper, we investigate several techniques for analyzing large-scale purchase and preference data for the purpose of producing useful recommendations to customers. In par- ticular, we apply a collection of algorithms such as tradi- tional data mining, nearest-neighbor collaborative filtering, and dimensionality reduction on two different data sets. The first data set was derived from the web-purchasing transac- tion of a large E-commerce company whereas the second data set was collected from MovieLens movie recommenda- tion site. For the experimental purpose, we divide the rec- ommendation generation process into three sub processes{ representation of input data, neighborhood formation, and recommendation generation. We devise different techniques for different sub processes and apply their combinations on our data sets to compare for recommendation quality and performance.
www.cs.umn.edu/~karypis/publications/Papers/PDF/ec00.pdf - reader comments
Added by James Thornton on 2001-02-01

Application of Dimensionality Reduction in Recommender System **(2000)**, *by Badrul M. Sarwar, George Karypis, Joseph A. Konstan, John T. Riedl (GroupLens Research Group Army HPC Research Center Department of Computer Science and Engineering University of Minnesota)*

We investigate the use of dimensionality reduction to improve performance for a new class of data analysis software called "recommender systems". Recommender systems apply knowledge discovery techniques to the problem of making product recommendations during a live customer interaction. These systems are achieving widespread success in E-commerce nowadays, especially with the advent of the Internet. The tremendous growth of customers and products poses three key challenges for recommender systems in the E-commerce domain. These are: producing high quality recommendations, performing many recommendations per second for millions of customers and products, and achieving high coverage in the face of data sparsity. One successful recommender system technology is collaborative filtering, which works by matching customer preferences to other customers in making recommendations. Collaborative filtering has been shown to produce high quality recommendations, but the performance degrades with the number of customers and products. New recommender system technologies are needed that can quickly produce high quality recommendations, even for very large-scale problems. This paper presents two different experiments where we have explored one technology called Singular Value Decomposition (SVD) to reduce the dimensionality of recommender system databases. Each experiment compares the quality of a recommender system using SVD with the quality of a recommender system using collaborative filtering. The first experiment compares the effectiveness of the two recommender systems at predicting consumer preferences based on a database of explicit ratings of products. The second experiment compares the effectiveness of the two recommender systems at producing Top-N lists based on a real-life customer purchase database from an E-Commerce site. Our experience suggests that SVD has the potential to meet many of the challenges of recommender systems, under certain conditions.
www.cs.umn.edu/~karypis/publications/Papers/PDF/webkdd.pdf - reader comments
Added by James Thornton on 2001-02-01

Artificial Ant Colonies in Digital Image Habitats - A Mass Behaviour Effect Study on Pattern Recognition **(2000)**, *by Vitorino Ramos (Technical Univ. of Lisbon, PORTUGAL), Filipe Almeida (VARIOGRAMA.com)*

(in, ANTS 2000 - 2nd International Workshop on Ant Algorithms - From Ant Colonies to Artificial Ants, Brussels, Belgium). Some recent studies have pointed that, the self-organization of neurons into brain-like structures, and the self-organization of ants into a swarm are similar in many respects. If possible to implement, these features could lead to important developments in pattern recognition systems, where perceptive capabilities can emerge and evolve from the interaction of many simple local rules. The principle of the method is inspired by the work of Chialvo and Millonas who developed the first numerical simulation in which swarm cognitive map formation could be explained. From this point, an extended model is presented in order to deal with digital image habitats, in which artificial ants could be able to react to the environment and perceive it. Evolution of pheromone fields point that artificial ant colonies could react and adapt appropriately to any type of digital habitat.
alfa.ist.utl.pt/~cvrm/staff/vramos/ref_29.html - reader comments
Added by Vitorino RAMOS on 2002-07-15

A User-Item Relevance Model for Log-based Collaborative Filtering (2006), *by Jun Wang (Delft University of Technology), Arjen P. de Vries (CWI),Marcel J.T. Reinders (Delft University of Technology), European Conference on Information Retrieval (ECIR 2006)*
> Implicit acquisition of user preferences makes log-based collaborative filtering favorable in practice to accomplish recommendations. In this paper, we follow a formal approach in text retrieval to re-formulate the problem. Based on the classic probability ranking principle, we propose a probabilistic user-item relevance model. Under this formal model, we show that user-based and item-based approaches are only two different factorizations with different independence assumptions. Moreover, we show that smoothing is an important aspect to estimate the parameters of the models due to data sparsity. By adding linear interpolation smoothing, the proposed model gives a probabilistic justification of using TFIDF-like item ranking in collaborative filtering. Besides giving the insight understanding of the problem of collaborative filtering, we also show experiments in which the proposed method provides a better recommendation performance on a music play-list data set.
> ict.ewi.tudelft.nl/pub/jun/ecir06.pdf - reader comments
> Added by Jun Wang on 2005-12-23

Automated Collaborative Filtering and Semantic Transports (**1997**), *by Alexander Chislenko*
> Automated Collaborative Filtering of information (ACF) is an unprecedented technology for distribution of opinions and ideas in society and facilitating contacts between people with similar interests. It automates and enhances existing mechanisms of knowledge distribution and dramatically increases their speed and efficiency. This allows to optimize knowledge flow in the society and accelerate the evolution of ideas in practically all subject areas. ACF also provides a superior tool for information retrieval systems that facilitates users' navigation in the sea of information in a meaningful and personalized way. This technology can be viewed as a semantic transport - a social utility that, after physical and data transports, transfers increasingly abstract and intelligent objects between previously isolated fragments of the social organism. As an artificial system that integrates and processes knowledge of multiple human participants, ACF represents an intermediate stage between human and purely artificial intelligence and lays the foundation for the future knowledge processing industry. This article discusses the premises and the historical analogs of ACF technology and suggests its possible uses as well as long-term economic and social implications.
> www.lucifer.com/~sasha/articles/ACF.html - reader comments
> Added by James Thornton on 2001-01-31

AuWeb-Collaborative Filtering: Recommending Music by Crawling The Web (**1999**), *by William W. Cohen and Wei Fan (AT&T Shannon Laboratories & Department of Computer Science, Columbia University)*
> We show that it is possible to collect data that is useful for collaborative filtering (CF) using an autonomous Web spider. In CF, entities are recommended to a new user based on the stated preferences of other, similar users. We describe a CF spider that collects from the Web lists of semantically related entities. These lists can then be used by existing CF algorithms by encoding them as "pseudo-users". Importantly, the spider can collect useful data without pre-programmed knowledge about the format of particular pages or particular sites. Instead, the CF spider uses commercial Web-search engines to find pages likely to contain lists in the domain of interest, and then applies previously-proposed heuristics [Cohen, 1999] to extract lists from these pages. We show that data collected by this spider is nearly as effective for CF as data collected from real users, and more effective than data collected by two plausible hand-programmed spiders. In some cases, autonomously spidered data can also be combined with actual user data to improve performance.
> www9.org/w9cdrom/266/266.html - reader comments
> Added by James Thornton on 2001-01-31

Beyond Document Similarity: Understanding Value-Based Search and Browsing Technologies (**0001**), *by Ungar, L. and D.P. Foster*
> In the face of small, one or two word queries, high volumes of diverse documents on the Web are overwhelming search and ranking technologies that are based on document similarity measures. The increase of multimedia data within documents sharply exacerbates the shortcomings of these approaches. Recently, research prototypes and commercial experiments have added techniques that augment similarity-based search and ranking. These techniques rely on judgments about the 'value' of documents. Judgments are obtained directly from users, are derived by conjecture based on observations of user behavior, or are surmised from analyses of documents and collections. All these systems have been pursued independently, and no common understanding of the underlying processes has been presented. We survey existing value-based approaches, develop a reference

architecture that helps compare the approaches, and categorize the constituent algorithms. We explain the options for collecting value metadata, and for using that metadata to improve search, ranking of results, and the enhancement of information browsing. Based on our survey and analysis, we then point to several open problems.
www-db.stanford.edu/pub/papers/info-filter.ps - reader comments
Added by James Thornton on 2001-01-31

Clustering Items for Collaborative Filtering (1999), by Mark O'Connor & Jon Herlocker (Dept. of Computer Science and Engineering, University of Minnesota)
This short paper reports on work in progress related to applying data partitioning/clustering algorithms to ratings data in collaborative filtering. We use existing data partitioning and clustering algorithms to partition the set of items based on user rating data. Predictions are then computed independently within each partition. Ideally, partitioning will improve the quality of collaborative filtering predictions and increase the scalability of collaborative filtering systems. We report preliminary results that suggest that partitioning algorithms can greatly increase scalability, but we have mixed results on improving accuracy. However, partitioning based on ratings data does result in more accurate predictions than random partitioning, and the results are similar to those when the data is partitioned based on a known content classification.
www.cs.umbc.edu/~ian/sigir99-rec/papers/oconner_m.pdf - reader comments
Added by James Thornton on 2001-01-31

CoFIND- an Experiment in N-dimensional Collaborative Filtering (1999), by Jon Dron, Richard Mitchell, Phil Siviter, Chris Boyne (Association for the Advancement of Computing in Education)
This paper reports on the development of CoFIND, a web-based n-dimensional collaborative filtering system that seeks to guide learners to relevant resources based upon not only the content of the resources but the qualities exhibited by those resources that make them useful learning material. Qualities provide the n-dimensions of this collaborative filter. Qualities and resources are generated collaboratively by the users of the system. CoFIND is designed to allow evolution to occur, which is discussed in the context of Darwinian theory and includes reference to current theories relating to the development of complex systems. The paper goes on to describe the implementation of the system and the results of an early pilot experiment involving a group of 42 students. It is concluded that, despite encouraging early results, some further work is needed to develop an effective interface and to embody the kind of complex interactions needed to generate spontaneous evolution.
www.it.bton.ac.uk/staff/jd29/ndim.html - reader comments
Added by James Thornton on 2001-01-31

Collaborative Filtering by Personality Diagnosis: A Hybird Memory-and-Model-Based Approach (2000), by David M. Pennock (NEC Research Institute), Eric Horvitz(Microsoft Research), Steve Lawrence (NEC Research Institute), and C.Lee Giles (Penn State University)
The growth of Internet commerce has stimulated the use of collaborative filtering (CF) algorithms as recommender systems. Such systems leverage knowledge about the known preferences of multiple users to recommend items of interest to other users. CF methods have been harnessed to make recommendations about such items as web pages, movies, books, and toys. Researchers have proposed and evaluated many approaches for generating recommendations. We describe and evaluate a new method called personality diagnosis (PD). Given a user's preferences for some items, we compute the probability that he or she is of the same "personality type" as other users, and, in turn, the probability that he or she will like new items. PD retains some of the advantages of traditional similarity-weighting techniques in that all data is brought to bear on each prediction and new data can be added easily and incrementally. Additionally, PD has a meaningful probabilistic interpretation, which may be leveraged to justify, explain, and augment results. We report empirical results on the EachMovie database of movie ratings, and on user profile data collected from the CiteSeer digital library of Computer Science research papers. The probabilistic framework naturally supports a variety of descriptive measurements -- in particular, we consider the applicability of a value of information (VOI) computation.
www.neci.nj.nec.com/homepages/dpennock/papers/pd-uai-00.ps - reader comments
Added by James Thornton on 2001-01-31

Collaborative Filtering by Personality Diagnosis: A Hybird Memroy-and-Model-Based Approach (1999), by Eric Horvitz (Microsoft Research)
The growth of Internet commerce has stimulated the use of collaborative filtering (CF) algorithms as

recommender systems.Such systems leverage knowledge about the known preferences of multiple users to recommend items of interest to other users. CF methods have been harnessed to make recommendations about such items as web pages, movies, books, and toys. Researchers have proposed many approaches for generating recommendations. We describe and evaluate a new method called personality diagnosis (PD). Given a user's preferences for some items, we compute the probability that he or she is of the same "personality type" as other users, and, in turn, the probability that he or she will like new items. PD retains some of the advantages of traditional similarity-weighting CF approaches in that all data is brought to bear on each prediction and new data can be added easily and incrementally. Additionally, PD has a meaningful probabilistic interpretation, which may be leveraged to justify, explain, and augment results. We show empirically that PD provides better predictions that all four of the algorithms tested by Breese et al. [1998] on the EachMovie database of movie ratings. The probabilistic framework naturally supports a variety of descriptive measurements— in particular, we briefly consider the applicability of a value of information (VOI) computation. www.research.microsoft.com/~horvitz/cfpd.htm - reader comments
Added by James Thornton on 2001-02-01

Collaborative filtering: Community values (**2000**), *by Karen H. Keeter (IBM)*
Collaborative filtering uses community opinion and behavior to determine the value of information and identify important trends. It has found application in targeted advertising, knowledge management and market segmentation. As processing speed and user base increase, the lack of social norms has become the last important barrier to widespread adoption and use. www.ibm.com/services/innovation/etrcollaborative_filtering.pdf - reader comments
Added by James Thornton on 2001-01-31

Collaborative Filtering with Privacy (**2002**), *by John Canny (Computer Science Division, UC Berkeley)*
Server-based collaborative filtering systems have been very successful in e-commerce and in direct recommenda-tion applications. In future, they have many potential ap-plications in ubiquitous computing settings. But today's schemes have problems such as loss of privacy, favoring retail monopolies, and with hampering diffusion of innova-tions. We propose an alternative model in which users con-trol all of their log data. We describe an algorithm whereby a community of users can compute a public "aggregate" of their data that does not expose individual users' data. The aggregate allows personalized recommendations to be computed by members of the community, or by outsiders. The numerical algorithm is fast, robust and accurate. Our method reduces the collaborative filtering task to an itera-tive calculation of the aggregate requiring only addition of vectors of user data. Then we use homomorphic encryption to allow sums of encrypted vectors to be computed and de-crypted without exposing individual data. We give verifica-tion schemes for all parties in the computation. Our system can be implemented with untrusted servers, or with addi-tional infrastructure, as a fully peer-to-peer (P2P) system. www.cs.berkeley.edu/~jfc/'mender/IEEESP02.pdf - reader comments
Added by James Thornton on 2003-02-09

Collaborative Filtering with Privacy via Factor Analysis (**2002**), *by John Canny (Computer Science Division, University of California Berkeley)*
Collaborative filtering is valuable in e-commerce, and for direct recommendations for music, movies, news etc. But today's systems use centralized databases and have several disadvantages, including privacy risks. As we move toward ubiquitous computing, there is a great potential for individ-uals to share all kinds of information about places and things to do, see and buy, but the privacy risks are severe. In this paper we introduce a peer-to-peer protocol for collaborative filtering which protects the privacy of individual data. A sec-ond contribution of this paper is a new collaborative filtering algorithm based on factor analysis which appears to be the most accurate method for CF to date. The new algorithm has other advantages in speed and storage over previous al-gorithms. It is based on a careful probabilistic model of user choice, and on a probabilistically sound approach to dealing with missing data. Our experiments on several test datasets show that the algorithm is more accurate than previously reported methods, and the improvements increase with the sparseness of the dataset. Finally, factor analysis with pri-vacy is applicable to other kinds of statistical analyses of survey or questionaire data scientists (e.g. web surveys or questionaires). www.cs.berkeley.edu/~jfc/'mender/sigir.pdf - reader comments
Added by James Thornton on 2003-02-09

Collaborative Interface Agents (**1998**), *by Yezdi Lashkari, Max Metral, and Pattie Maes (M.I.T. Media Lab)*

Interface agents are semi-intelligent systems which assist users with daily computer-based tasks. Recently, various researchers have proposed a learning approach towards building such agents and some working prototypes have been demonstrated. Such agents learn by `watching over the shoulder' of the user and detecting patterns and regularities in the user's behavior. Despite the successes booked, a major problem with the learning approach is that the agent has to learn from scratch and thus takes some time becoming useful. Secondly, the agent's competence is necessarily imited to actions it has seen the user perform. Collaboration between agents assisting different users can alleviate both of these problems. We present a framework for multi-agent collaboration and discuss results of a working prototype, based on learning agents for electronic mail.
mevard.www.media.mit.edu/groups/agents/publications/aaai-ymp/aaai.html - reader comments
Added by James Thornton on 2001-01-31

Collaborative value filtering on the Web (1998), by Gerard RodrÃguez-MulÃ , Hector GarcÃa-Molina and Andreas Paepcke (Digital Libraries Lab [InfoLab], Stanford University)
This paper presents a prototype (KSS) that monitors the behavior of a community of users for collaborative filtering and community-based navigation purposes. Our hope is to develop mechanisms for sharing browsing expertise and better understand their access patterns. The KSS architecture is based on a federation of KSS proxies.
www7.scu.edu.au/programme/posters/1851/com1851.htm - reader comments
Added by James Thornton on 2001-01-31

Combining Content-Based and Collaborative Filters in an Online Newspaper (1999), by Mark Claypool Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes and Matthew Sartin (ACM SIGIR Workshop on Recommender Systems Berkeley, CA)
The explosive growth of mailing lists, Web sites and Usenet news demands effective filtering solutions. Collaborative filtering combines the informed opinions of humans to make personalized, accurate predictions. Content-based filtering uses the speed of computers to make complete, fast predictions. In this work, we present a new filtering approach that combines the coverage and speed of content-filters with the depth of collaborative filtering. We apply our research approach to an online newspaper, an as yet untapped opportunity for filters useful to the wide-spread news reading populace. We present the design of our filtering system and describe the results from preliminary experiments that suggest merits to our approach.
www.cs.wpi.edu/~claypool/papers/content-collab/content-collab.pdf - reader comments
Added by James Thornton on 2001-02-01

Community-Based Ratings for the Net (1995), by Alan Wexelblat, Lenny Foner, Rich Lethin, James O'Toole, Yezdi Lashkari, Brian Behlendorf (M.I.T. Media Lab)
This document web lays out an alternative to current proposals for standards and ratings of World Wide Web documents. The objective of this proposal is to explain how, using current technology, groups of like-minded people can work together to provide more flexible, more personalized, and more comprehensive information about net resources.
mevard.www.media.mit.edu/people/wex/rate-proposal-head.html - reader comments
Added by James Thornton on 2001-01-31

Content-based Collaborative Information Filtering: Actively Learning to Classify and Recommend Documents (0001), by Joaquin Delgado, Naohiro Ishii, and Tomoki Ura (Department of Intelligence & Computer Science Nagoya Institute of Technology)
Next generation of intelligent information systems will rely on cooperative agents for playing a fundamental role in actively searching and finding relevant information on behalf of their users in complex and open environments, such as the Internet. Whereas relevant can be defined solely for a specific user, and under the context of a particular domain or topic. On the other hand shared "social" information can be used to improve the task of retrieving relevant information, and for refining each agent's particular knowledge. In this paper, we combine both approaches developing a new content-based filtering technique for learning up-to-date users' profile that serves as basis for a novel collaborative information-filtering algorithm. We demonstrate our approach through a system called RAAP (Research Assistant Agent Project) devoted to support collaborative research by classifying domain specific information, retrieved from the Web, and recommending these "bookmarks" to other researcher with similar research interests.
citeseer.nj.nec.com/delgado98intelligent.html - reader comments
Added by James Thornton on 2001-02-01

Content Filtering Technologies and Internet Service Providers: Enabling User Choice **(2000)**, *by Michael Sheperd and Carolyn Watters (Faculty of Computer Science, Dalhousie University)*
>   This project investigates the set of mechanisms that Internet Service Providers (ISPs) have the option to provide and that users can choose to utilize in order to filter the content delivered to users over the Internet and to allow authorized access to that content. The report is purely descriptive of the mechanisms available and does not provide policy or legal advice or recommendations.
>   www.cs.dal.ca/~shepherd/filtering/ISPweb.htm - reader comments
>   Added by James Thornton on 2001-01-31

Dependency Networks for Inference, Collaborative Filtering, and Data Visualization **(2000)**, *by David Heckerman, David Maxwell Chickering, Christopher Meek, Robert Rounthwaite, and Carl Kadi (Microsoft Research)*
>   We describe a graphical model for probabilistic relationships--an alternative to the Bayesian network--called a dependency network. The graph of a dependency network, unlike a Bayesian network, is potentially cyclic. The probability component of a dependency network, like a Bayesian network, is a set of conditional distributions, one for each node given its parents. We identify several basic properties of this representation and describe a computationally efficient procedure for learning the graph and probability components from data. We describe the application of this representation to probabilistic inference, collaborative filtering (the task of predicting preferences), and the visualization of acausal predictive relationships.
>   www.ai.mit.edu/projects/jmlr/papers/volume1/heckerman00a/heckerman00a.pdf - reader comments
>   Added by James Thornton on 2001-02-26

Eigentaste: A Constant Time Collaborative Filtering Algorithm **(2000)**, *by Eigentaste: Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins (IEOR and EECS Departments University of California, Berkeley)*
>   Eigentaste is a collaborative filtering algorithm that uses _universal queries_ to elicit real-valued user ratings on a common set of items and applies principal component analysis (PCA) to the resulting dense subset of the ratings matrix. PCA facilitates dimensionality reduction for offline clustering of users and rapid computation of recommendations. For a database of n users, standard nearest-neighbor techniques require O(n) processing time to compute recommendations, whereas Eigentaste requires O(1) (constant) time. We compare Eigentaste to alternative algorithms using data from _Jester_, an online joke recommending system. Jester has collected approximately 2,500,000 ratings from 57,000 users. We use the Normalized Mean Absolute Error (NMAE) measure to compare performance of different algorithms. In the Appendix we use Uniform and Normal distribution models to derive analytic estimates of NMAE when predictions are random. On the Jester dataset, Eigentaste computes recommendations two orders of magnitude faster with no loss of accuracy. (The Jester dataset including ratings from approximately 18,000 anonymous users is available by request: contact goldberg@ieor.berkeley.edu with contact information and a description of intended research.)
>   www.ieor.berkeley.edu/~goldberg/pubs/eigentaste.pdf - reader comments
>   Added by James Thornton on 2001-02-01

Empirical Analysis of Predictive Algorithms for Collaborative Filtering **(1998)**, *by by Jack Breese, David Heckerman, and Carl Kadie (Microsoft Research)*
>   Collaborative filtering or recommender systems use a database about user preferences to predict additional topics or products a new user might like. In this paper we describe several algorithms designed for this task, including techniques based on correlation coefficients, vector-based similarity calculations, and statistical Bayesian methods. We compare the predictive accuracy of the various methods in a set of representative problem domains. We use two basic classes of evaluation metrics. The first characterizes accuracy over a set of individual predictions in terms of average absolute deviation. The second estimates the utility of a ranked list of suggested items. This metric uses an estimate of the probability that a user will see a recommendation in an ordered list. Experiments were run for datasets associated with 3 application areas, 4 experimental protocols, and the 2 evaluation metrics for the various algorithms. Results indicate that for a wide range of conditions, Bayesian networks with decision trees at each node and correlation methods outperform Bayesian-clustering and vector-similarity methods. Between correlation and Bayesian networks, the preferred method depends on the nature of the dataset, nature of the application (ranked versus one-by-one presentation), and the availability of votes with which to make predictions. Other considerations include the size of database, speed of predictions, and learning time.
>   www.research.microsoft.com/users/breese/cfalgs.html - reader comments
>   Added by James Thornton on 2001-01-31

Evaluation of Item-Based Top-N Recommendation Algorithms (**2000**), *by George Karypis (University of Minnesota, Department of Computer Science / Army HPC Research Center)*

> The explosive growth of the world-wide-web and the emergence of e-commerce has led to the development of recommender systems- a personalized information filtering technology used to identify a set of N items that will be of interest to a certain user. User-based Collaborative filtering is the most successful technology for building recommender systems to date, and is extensively used in many commercial recommender systems. Unfortunately, the computational complexity of these methods grows linearly with the number of customers that in typical commercial applications can grow to be several millions. To address these scalability concerns item-based recommendation techniques have been developed that analyze the user-item matrix to identify relations between the different items, and use these relations to compute the list of recommendations. In this paper we present one such class of item-based recommendation algorithms that first determine the similari-ties between the various items and then used them to identify the set of items to be recommended. The key steps in this class of algorithms are (i) the method used to compute the similarity between the items, and (ii) the method used to combine these similarities in order to compute the similarity between a basket of items and a candidate recommender item. Our experimental evaluation on five different datasets show that the proposed item-based algorithms are up to 28 times faster than the traditional user-neighborhood based recommender systems and provide recommendations whose quality is up to 27% better.
> www-users.cs.umn.edu/~karypis/publications/Papers/PDF/itemrs.pdf - reader comments
> Added by James Thornton on 2001-02-01

Evolving a Stigmergic Self-Organized Data-Mining (**2004**), *by Vitorino Ramos (CVRM-IST, Technical Univ. of Lisbon, PORTUGAL), Ajith Abraham (Oklahoma Univ., USA)*

> Self-organizing complex systems typically are comprised of a large number of frequently similar components or events. Through their process, a pattern at the global-level of a system emerges solely from numerous interactions among the lower-level components of the system. Moreover, the rules specifying interactions among the systemÂ's components are executed using only local information, without reference to the global pattern, which, as in many real-world problems is not easily accessible or possible to be found. Stigmergy, a kind of indirect communication and learning by the environment found in social insects is a well know example of self-organization, providing not only vital clues in order to understand how the components can interact to produce a complex pattern, as can pinpoint simple biological non-linear rules and methods to achieve improved artificial intelligent adaptive categorization systems, critical for Data-Mining. On the present work it is our intention to show that a new type of Data-Mining can be designed based on Stigmergic paradigms, taking profit of several natural features of this phenomenon. By hybridizing bio-inspired Swarm Intelligence with Evolutionary Computation we seek for an entire distributed, adaptive, collective and cooperative self-organized Data-Mining. As a real-world / real-time test bed for our proposal, World-Wide-Web Mining will be used. Having that purpose in mind, Web usage Data was collected from the Monash UniversityÂ's Web site (Australia), with over 7 million hits every week. Results are compared to other recent systems, showing that the system presented is by far promising.
> alfa.ist.utl.pt/~cvrm/staff/vramos/ref_50.html - reader comments
> Added by Vitorino RAMOS on 2004-01-22

GroupLens: an open architecture for collaborative filtering of netnews (**1994**), *by Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl*

> Collaborative filters help people make choices based on the opinions of other people. GroupLens is a system for collaborative filtering of netnews, to help people find articles they will like in the huge stream of available articles. News reader clients display predicted scores and make it easy for users to rate articles after they read them. Rating servers, called Better Bit Bureaus, gather and disseminate the ratings. The rating servers predict scores based on the heuristic that people who agreed in the past will probably agree again. Users can protect their privacy by entering ratings under pseudonyms, without reducing the effectiveness of the score prediction. The entire architecture is open: alternative software for news clients and Better Bit Bureaus can be developed independently and can interoperate with the components we have developed.
> www.acm.org/pubs/citations/proceedings/cscw/192844/p175-resnick/ - reader comments
> Added by James Thornton on 2001-01-31

Implicit Interest Indicators (**2000**), *by Mark Claypool, Phong Le, Makoto Waseda and David Brown (Worcester Polytechnic Institute)*

> Recommender systems provide personalized suggestions about items that users will find interesting.

Typically, recommender systems require a user interface that can ``intelligently'' determine the interest of a user and use this information to make suggestions. The common solution, ``explicit ratings'', where users tell the system what they think about a piece of information, is well-understood and fairly precise. However, having to stop to enter explicit ratings can alter normal patterns of browsing and reading. A more ``intelligent'' method is to use implicit ratings, where a rating is obtained by a method other than obtaining it directly from the user. These implicit interest indicators have obvious advantages, including removing the cost of the user rating, and that every user interaction with the system can contribute to an implicit rating. Current recommender systems mostly do not use implicit ratings, nor is the ability of implicit ratings to predict actual user interest well-understood. This research studies the correlation between various implicit ratings and the explicit rating for a single Web page. A Web browser was developed to record the user's actions (implicit ratings) and the explicit rating of a page. Actions included mouse clicks, mouse movement, scrolling and elapsed time. This browser was used by over 80 people that browsed more than 2500 Web pages. Using the data collected by the browser, the individual implicit ratings and some combinations of implicit ratings were analyzed and compared with the explicit rating. We found that the time spent on a page, the amount of scrolling on a page and the combination of time and scrolling had a strong correlation with explicit interest, while individual scrolling methods and mouse-clicks were ineffective in predicting explicit interest.
www.cs.wpi.edu/~claypool/papers/iii/ - reader comments
Added by Mark Claypool on 2001-06-13

Implicit Rating and Filtering (**1997**), *by David M. Nichols (Computing Department, Lancaster University)*
Social filtering systems that use explicit ratings require a large number of ratings to remain viable. The effort involved for a user to rate a document may outweigh any benefit received, leading to a shortage of ratings. One approach to this problem is to use implicit ratings: where user actions are recorded and a rating is inferred from the recorded data. This paper discusses the costs and benefits of using implicit ratings for information filtering applications.
www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/delos5.html - reader comments
Added by James Thornton on 2001-01-31

Item-based Collaborative Filtering Recommendation Algorithms (**2001**), *by Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl (GroupLens Research Group/Army HPC Research Center Department of Computer Science and Engineering University of Minnesota)*
Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations for information, products or services during a live interaction. These systems, especially the k-nearest neighbor collaborative filtering based ones, are achieving widespread success on the Web. The tremendous growth in the amount of available information and the number of visitors to Web sites in recent years poses some key challenges for recommender systems. These are: producing high quality recommendations, performing many recommendations per second for millions of users and items and achieving high coverage in the face of data sparsity. In traditional collaborative filtering systems the amount of work increases with the number of participants in the system. New recommender system technologies are needed that can quickly produce high quality recommendations, even for very large-scale problems. To address these issues we have explored item-based collaborative filtering techniques. Item-based techniques first analyze the user-item matrix to identify relationships between different items, and then use these relationships to indirectly compute recommendations for users.
www.cs.umn.edu/Research/GroupLens/papers/pdf/www10_sarwar.pdf - reader comments
Added by James Thornton on 2003-02-09

Learning the Structure of Utility Graphs Used in Multi-Issue Negotiation through Collaborative Filtering (**2005**), *by Valentin Robu, J.A. La Poutre (CWI, Dutch National Research Center for Mathematics and Computer Science, Amsterdam). Presented at PRIMA 2005 conference, Kuala Lumpur, Malaysia.*
We study the problem of automating complex, multi-issue negotiations between electronic merchants and a buyers in an e-commerce setting. Utility graphs have been shown to be a powerful formalism to model buyer preferences in such situations, especially if there are non-linearity (i.e. complementarity/ subsitutability) effects between items on sale. This paper proposes a method for constructing the utility graphs of buyers auto- matically, based on previous sales data. Our method is based on techniques inspired from item-based collaborative filtering. Experimental results show that our approach is able to retrieve the structure of utility graphs online, with a high degree of accuracy. This

enables agents to reach efficient outcomes during the negotiation, even if the utility function of the buyer is not fully elicited during the negotiation (a process which can be very costly).
homepages.cwi.nl/~robu/prima2005.pdf - reader comments
Added by Valentin Robu on 2005-11-03

Multi-Agent Learning in Recommender Systems for Information Filtering on the Internet (**2001**), *by Joaquin Delgado (TripleHop Technologies, Inc.), Naohiro Ishii (Nagoya Institute of Technology)*
Recommender Systems (RS), allow users to share information about items they like or dislike and obtain, in a timely fashion, recommendations based on predictions about unseen items (physical or information goods and/or services). In this process, users' preferences are considered to be the learning target functions. We study Agent-based Recommender Systems (ARS) under the scope of online learning in Multi-Agent systems (MAS). This approach models the problem as a pool of independent cooperative predictor agents, one per each user (the masters) in the system, in situations in which each agent (the learners) faces a sequence of trials, with a prediction to make in every step, eventually getting the correct value from its master. Each learner is willing to discover the degree of similarity among the target function of its master and those of other agents' masters (i.e. preference similarity). The agent uses this information for the calculation of its own prediction task, the goal being to make as few mistakes as possible. A simple, yet effective method is introduced in order to construct a compound algorithm for each agent by combining memory-based individual prediction and online weighted-majority voting. We give a theoretical mistake bound for this algorithm that is closely related to the total loss of the best predictor agent in the pool. Finally, we conduct some experiments obtaining results that empirically support these ideas and theories.
International Journal of Cooperative Information Systems Vol. 10, Nos. 1 & 2 (2001) 81-100
Copyright 2001 World Scientific Publishing Company
www.triplehop.com/research/jdelgado-cis.pdf - reader comments
Added by Joaquin Delgado on 2001-04-11

Pointing the Way: Active Collaborative Filtering (**1995**), *by David Maltz and Kate Ehrlich (Carnegie Mellon Univeristy and Lotus)*
Collaborative filtering is based on the premise that people looking for information should be able to make use of what others have already found and evaluated. Current collaborative filtering systems provide tools for readers to filter documents based on which ones were read and liked by previous readers. This paper describes a different type of collaborative filtering system in which people who find interesting documents actively send "pointers" to those documents to their colleagues. A "pointer" contains a hypertext link to the source document as well as contextual information intended to help the recipient determine the potential interest and relevance of the document prior to accessing it. A preliminary version of our system has already proven easy to use, with people using it to "bookmark" documents, send pointers to their colleagues and create "digests" that combine pointers with original text. Based on our experience we discuss the benefits of this form of filtering as well as its limitations.
www.cs.cmu.edu/~dmaltz/ACF95-draft8.txt - reader comments
Added by James Thornton on 2001-01-31

RACOFI: A Rule-Applying Collaborative Filtering System (**2003**), *by Michelle Anderson, Marcel Ball, Harold Boley, Stephen Greene, Nancy Howse, Daniel Lemire, Sean McGrath (National Research Council of Canada)*
In this paper we give an overview of the RACOFI (Rule-Applying Collaborative Filtering) multidimensional rating system and its related technologies. This will be exemplified with RACOFI Music, an implemented collaboration agent that assists on-line users in the rating and recommendation of audio (Learning) Objects. It lets users rate contemporary Canadian music in the five dimensions of impression, lyrics, music, originality, and production. The collaborative filtering algorithms STI Pearson, STIN2, and the Per Item Average algorithms are then employed together with RuleML-based rules to recommend music objects that best match user queries. RACOFI has been on-line since August 2003 at http://racofi.elg.ca. .
www.ondelette.com/lemire/abstracts/COLA2003.html - reader comments
Added by Daniel Lemire on 2003-09-18

Scale And Translation Invariant Collaborative Filtering Systems (**2003**), *by Daniel Lemire*
Collaborative filtering systems are prediction algorithms over sparse data sets of user preferences. We modify a wide range of state-of-the-art collaborative filtering systems to make them scale and translation invariant and generally improve their accuracy without increasing their computational cost.

Using the EachMovie and the Jester data sets, we show that learning-free constant time scale and translation invariant schemes outperforms other learning-free constant time schemes by at least 3% and perform as well as expensive memory-based schemes (within 4%). Over the Jester data set, we show that a scale and translation invariant Eigentaste algorithm outperforms Eigentaste 2.0 by 20%. These results suggest that scale and translation invariance is a desirable property.
www.ondelette.com/lemire/abstracts/IR2003.html - reader comments
Added by Daniel Lemire on 2003-10-21

Self-Organized Stigmergic Document Maps: Environment as a Mechanism for Context Learning (2002), by Vitorino Ramos (Technical Univ. of Lisbon, PORTUGAL), Juan J. Merelo (Granada Univ., SPAIN)
(in, MAEB 2002 - 1st Spanish Conference on Evolutionary and Bio-Inspired Algorithms, Merida, Spain). Social insect societies and more specifically ant colonies, are distributed systems that, in spite of the simplicity of their individuals, present a highly structured social organization. As a result of this organization, ant colonies can accomplish complex tasks that in some cases exceed the individual capabilities of a single ant. The study of ant colonies behavior and of their self-organizing capabilities is of interest to knowledge retrieval/management and decision support systems sciences, because it provides models of distributed adaptive organization which are useful to solve difficult optimization, classification, and distributed control problems, among others. In the present work we overview some models derived from the observation of real ants, emphasizing the role played by stigmergy as distributed communication paradigm, and we present a novel strategy to tackle unsupervised clustering as well as data retrieval problems. The present ant clustering system (ACLUSTER) avoids not only short-term memory based strategies, as well as the use of several artificial ant types (using different speeds), present in some recent approaches. Moreover and according to our knowledge, this is also the first application of ant systems into textual document clustering.
alfa.ist.utl.pt/~cvrm/staff/vramos/ref_42.html - reader comments
Added by Vitorino RAMOS on 2002-07-15

Semantic Ratings and Heuristic Similarity for Collaborative Filtering (2000), by Robin Burkey (Department of Information and Computer Science, University of California, Irvine)
Collaborative filtering systems make recommendations based on ratings of user preference. Usually, the ratings are uni-dimensional (e.g. like vs. dislike), and can be either explicitly elicited from users or, more typically, are implicitly generated from observations of user behavior. This research examines multi-dimensional or semantic ratings in which a system gets information about the reason behind a preference. Such multi-dimensional ratings can be projected onto a single dimension, but experiments show that metrics in which the semantic meaning of each rating is taken into account have markedly superior performance.
www.igec.umbc.edu/kbem/final/burke.pdf - reader comments
Added by James Thornton on 2001-01-31

Slope One Predictors for Online Rating-Based Collaborative Filtering (2005), by Daniel Lemire, Anna Maclachlan
Rating-based collaborative filtering is the process of predicting how a user would rate a given item from other user ratings. We propose three related slope one schemes with predictors of the form f(x) = x + b, which precompute the average difference between the ratings of one item and another for users who rated both. Slope one algorithms are easy to implement, efficient to query, reasonably accurate, and they support both online queries and dynamic updates, which makes them good candidates for real-world systems. The basic slope one scheme is suggested as a new reference scheme for collaborative filtering. By factoring in items that a user liked separately from items that a user disliked, we achieve results competitive with slower memory-based schemes over the standard benchmark EachMovie and Movielens data sets while better fulfilling the desiderata of CF applications.
www.ondelette.com/lemire/documents/publications/racofi_nrc.pdf - reader comments
Added by Daniel Lemire on 2005-01-09

Social Information Filtering: Algorithms for Automating "Word of Mouth" (1995), by Upendra Shardanand and Pattie Maes (MIT Media-Lab)
This paper describes a technique for making personalized recommendations from any type of database to a user based on similarities between the interest profile of that user and those of other users. In particular, we discuss the implementation of a networked system called Ringo, which makes personalized recommendations for music albums and artists. Ringo's database of users and artists grows dynamically as more people use the system and enter more information. Four different

algorithms for making recommendations by using social information filtering were tested and compared. We present quantitative and qualitative results obtained from the use of Ringo by more than 2000 people.
www.acm.org/sigchi/chi95/Electronic/documnts/papers/us_bdy.htm - reader comments
Added by James Thornton on 2001-02-28

Swarms on Continuous Data (**2003**), *by Vitorino Ramos (Technical Univ. of Lisbon, Portugal), Ajith Abraham (Olklahoma Univ., USA)*

[in CEC03 - Congress on Evolutionary Computation, IEEE Press, Canberra, Australia, 8-12 Dec. 2003] While being it extremely important, many Exploratory Data Analysis (EDA) systems have the inhability to perform classification and visualization in a continuous basis or to self-organize new data-items into the older ones (evenmore into new labels if necessary), which can be crucial in KDD - Knowledge Discovery, Retrieval and Data Mining Systems (interactive and online forms of Web Applications are just one example). This disadvantge is also present in more recent approaches using Self-Organizing Maps. On the present work, and exploiting past sucesses in recently proposed Stigmergic Ant Systems a robust online classifier is presented, which produces class decisions on a continuous stream data, allowing for continuous mappings. Results show that increasingly better results are achieved, as demonstraded by other authors in different areas.
alfa.ist.utl.pt/~cvrm/staff/vramos/Vramos-CEC03a.pdf - reader comments
Added by Vitorino RAMOS on 2003-09-20

The Anatomy of a Large-Scale Hypertextual Web Search Engine (**1998**), *by Sergey Brin and Lawrence Page (Computer Science Department, Stanford University)*

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu. To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want.
www-db.stanford.edu/~backrub/google.html - reader comments
Added by James Thornton on 2001-01-31

The Effects of Singular Value Decomposition on Collaborative Filtering (**1998**), *by Michael H. Pryor (Dartmouth College)*

As the information on the web increases exponentially, so do the efforts to automatically filter out useless content and to search for interesting content. Through both explicit and implicit actions, users define where their interests lie. Recent efforts have tried to group similar users together in order to better use this data to provide the best overall filtering capabilities to everyone. This thesis discusses ways in which linear algebra, specifically the singular value decomposition, can be used to augment these filtering capabilities to provide better user feedback. The goal is to modify the way users are compared with one another, so that we can more efficiently predict similar users. Using data collected from the PhDs.org website, we tested our hypothesis on both explicit web page ratings and implicit visits data
www.cs.dartmouth.edu/reports/abstracts/TR98-338/ - reader comments
Added by James Thornton on 2001-01-31

The Hidden Web (**1997**), *by Henry Kautz, Bart Selman, and Mehul Shah (The American Association for Artificial Intelligence)*

The difficulty of finding information on the World Wide Web by browsing hypertext documents has led to the development and deployment of various search engines and indexing techniques. However,

many information-gathering tasks are better handled by finding a referral to a human expert rather than by simply interacting with online information sources. A personal referral allows a user to judge the quality of the information he or she is receiving as well as to potentially obtain information that is deliberately not made public. The process of finding an expert who is both reliable and likely to respond to the user can be viewed as a search through the network of social relationships between individuals as opposed to a search through the network of hypertext documents. The goal of the REFERRAL WEB Project is to create models of social networks by data mining the web and develop tools that use the models to assist in locating experts and related information search and evaluation tasks.
www.cs.washington.edu/homes/kautz/papers/aimag.pdf - reader comments
Added by James Thornton on 2001-01-31

The MC2 Project [Machines of Collective Conscience] **(2001)**, *by Vitorino Ramos (Technical Univ. of Lisbon, PORTUGAL)*
>(in, the Official Newspaper of the UTOPIA Biennial Art Exposition, Cascais, Portugal). Imagine a Å"machineÅ" where there is no pre-commitment to any particular representational scheme: the desired behaviour is distributed and roughly specified simultaneously among many parts, but there is minimal specification of the mechanism required to generate that behaviour, i.e. the global behaviour evolves from the many relations of multiple simple behaviours. A machine that lives to and from/with Synergy. We believe that these are the first steps into the design of truly collective, flexible, cognitive and adaptive forms of information structures, whatever they may be, or whatever they may represent, among many possible and specific contexts.
>alfa.ist.utl.pt/~cvrm/staff/vramos/ref_36.html - reader comments
>Added by Vitorino RAMOS on 2002-07-15

Trust-aware Collaborative Filtering for Recommender Systems **(2004)**, *by Paolo Massa (ITC/iRST - Trento - Italy), Paolo Avesani (ITC/iRST - Trento - Italy)*
>Recommender Systems allow people to nd the resources they need by making use of the experiences and opinions of their nearest neigh- bours. Costly annotations by experts are replaced by a distributed pro- cess where the users take the initiative. While the collaborative approach enables the collection of a vast amount of data, a new issue arises: the quality assessment. The elicitation of trust values among users, termed "web of trust", allows a twofold enhancement of Recommender Systems. Firstly, the ltering process can be informed by the reputation of users which can be computed by propagating trust. Secondly, the trust metrics can help to solve a problem associated with the usual method of simi- larity assessment, its reduced computability. An empirical evaluation on Epinions.com dataset shows that trust propagation allows to increase the coverage of Recommender Systems while preserving the quality of pre- dictions. The greatest improuvements are achieved for new users, who provided few ratings.
>sra.itc.it/people/massa/publications/massa_paolo_coopis_2004_trust-aware_Collaborative_Filtering_for_Recommender_Systems.pdf - reader comments
>Added by paolo massa on 2005-04-05

Trust in Recommender Systems **(2005)**, *by John OÂ'Donovan, Barry Smyth (Adaptive Information Cluster School of Computer Science and Informatics University College Dublin Belfield, Dublin 4 Ireland)*
>Recommender systems have proven to be an important response to the information overload problem, by providing users with more proactive and personalized information services. And collaborative filtering techniques have proven to be an vital component of many such recommender systems as they facilitate the generation of high-quality recommendations by leveraging the preferences of communities of similar users. In this paper we suggest that the traditional emphasis on user similarity may be overstated. We argue that additional factors have an important role to play in guiding recommendation. Specifically we propose that the trustworthiness of users must be an important consideration. We present two computational models of trust and show how they can be readily incorporated into standard collaborative filtering frameworks in a variety of ways. We also show how these trust models can lead to improved predictive accuracy during recommendation.
>delivery.acm.org/10.1145/1050000/1040870/p167-odonovan.pdf?
>key1=1040870&key2=9333540311&coll=GUIDE&dl=GUIDE&CFID=58637181&CFTOKEN=67159970
>- reader comments
>Added by John o'donovan on 2005-10-27

Using a Semantic User Model to Filter the World Wide Web Proactively **(1997)**, *by Joep Simons (University*

of Nijmegen, Netherlands)

The research in this paper aims at using world knowledge to aid the user in retrieving information from the World Wide Web. Some issues are identified together with methods to address them. 1 Introduction Information retrieval systems are consulted to meet an information need. First, users must translate their internal representation of the information need to a query the system understands. Second, the system must match the queries of the users with the stored characterizations of the documents in a fixed archive. An information filtering system deals with a user's information need that is relatively stable over time. This is represented by a user profile. A profile is used to filter a rapidly changing archive by viewing it as a stream of documents. Finally, a proactive filter.
www.cs.usask.ca/um-inc/um_97/gz/SimonsJ.ps.gz - reader comments
Added by James Thornton on 2001-01-31

Using Mixture Models for Collaborative Filtering **(2004)**, *by Jon Kleinberg, Mark Sandler (Cornell University)*

A {it collaborative filtering system} at an e-commerce site or similar service uses data about aggregate user behavior to make recommendations tailored to specific user interests. We develop recommendation algorithms with provable performance guarantees in a probabilistic {it mixture model} for collaborative filtering proposed by Hoffman and Puzicha. We identify certain novel parameters of mixture models that are closely connected with the best achievable performance of a recommendation algorithm; we show that for any system in which these parameters are bounded, it is possible to give recommendations whose quality converges to optimal as the amount of data grows. All our bounds depend on a new measure of independence that can be viewed as an L_1-analogue of the smallest singular value of a matrix. Using this, we introduce a technique based on generalized pseudoinverse matrices and linear programming for handling sets of high-dimensional vectors. We also show that standard approaches based on L_2 spectral methods are not strong enough to yield comparable results, thereby suggesting some inherent limitations of spectral analysis.
www.cs.cornell.edu/~sandler/mmicf.ps - reader comments
Added by Mark Sandler on 2005-11-29

Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming **(2003)**, *by Ajith Abraham (Oklahoma Univ., USA), Vitorino Ramos (Technical Univ. of Lisbon, Portugal)*

[in CEC03 - Congress on Evolutionary Computation, IEEE Press, Canberra, Australia, 8-12 Dec. 2003] The rapid e-commerce growth has made both business community and customers face a new situation. Due to intense competition on one hand and the customer's option to choose from several alternatives business community has realized the necessity of intelligent marketing strategies and relationship management. Web usage mining attempts to discover useful knowledge from the secondary data obtained from the interactions of the users with the Web. Web usage mining has become very critical for effective Web site management, creating adaptive Web sites, business and support services, personalization, network traffic flow analysis and so on. The study of ant colonies behavior and their self-organizing capabilities is of interest to knowledge retrieval/management and decision support systems sciences, because it provides models of distributed adaptive organization, which are useful to solve difficult optimization, classification, and distributed control problems, among others. In this paper, we propose an ant clustering algorithm to discover Web usage patterns (data clusters) and a linear genetic programming approach to analyze the visitor trends. Empirical results clearly shows that ant colony clustering performs well when compared to a self-organizing map (for clustering Web usage patterns) even though the performance accuracy is not that efficient when comparared to evolutionary-fuzzy clustering (i-miner) approach.
alfa.ist.utl.pt/~cvrm/staff/vramos/Vramos-CEC03b.pdf - reader comments
Added by Vitorino RAMOS on 2003-09-20

Wide Area Collaboration: A Proposed Application **(1997)**, *by John Caron (University of Colorado, Boulder)*

In this paper I explore an idea for a Web-based wide-area collaborative application for capturing and structuring knowledge. Wide-area collaboration requires that interactions be asynchronous, and that will be the focus in this paper. Part two presents further motivation and context. Part three contains summaries of related work. Part four presents elements of the application requirements and high level design, including proposed user roles. Part five has some modest contributions to engineering and implementation issues, and part six summarizes and concludes with thoughts on future work.
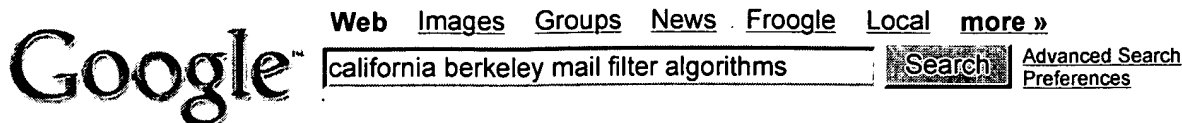www.unidata.ucar.edu/staff/caron/collab/wa_collab.html - reader comments
Added by James Thornton on 2001-01-31

<u>Chubbuck Mortgage Refinance</u> (rates are still low)
cofedbank.com Incredible Home Loans, No Hassles

James Thornton,<<u>james@jamesthornton.com</u>> Services: <u>Software Developer</u>	Electric Speed: <u>Computer Internet And Technology Consultant Company</u>

Web   Images   Groups   News   Froogle   Local   **more »**

Google |california berkeley mail filter algorithms | Search   Advanced Search
                                                                        Preferences

Web   Results **1 - 10** of about **343,000** for <u>california</u> <u>berkeley</u> <u>mail</u> <u>filter</u> <u>algorithms</u>. **(0.12** seconds)

Welcome to Feng Zhou's homepage
University of **California, Berkeley** 445 Soda Hall Work: 510-643-9435 ... The 5th
International Conference on **Algorithms** and Architecture for Parallel ...
www.cs.**berkeley**.edu/~zf/ - 13k - <u>Cached</u> - <u>Similar pages</u>

CS 170: Efficient **Algorithms** and Intractable Problems
If your question is personal or not of interest to other students, you may send
email to cs170@cory.eecs.**berkeley**.edu. **Email** to cs170@cory is forwarded to ...
www.cs.**berkeley**.edu/~jordan/courses/170-fall05/ - 30k - <u>Cached</u> - <u>Similar pages</u>
[<u>More results from www.cs.berkeley.edu</u>]

Citations: A New Linear-Time Collaborative **Filtering Algorithm** ...
... which has been developing at University of **California, Berkeley** [4]. ...
Jester 2.0: A new lineartime collaborative **filtering algorithm** applied to jokes ...
citeseer.ist.psu.edu/context/948327/0 - 15k - <u>Cached</u> - <u>Similar pages</u>

Breaking a Visual CAPTCHA
These companies offer free **email** accounts, intended for use by humans. ...
Our current **algorithm** can find 3 correct words and pass this Gimpy test 33% of ...
www.cs.sfu.ca/~mori/research/gimpy/ - 9k - <u>Cached</u> - <u>Similar pages</u>

exploring enron | v1::white
University of **California, Berkeley** Fall Semester, 2004 Click any image to get a
full-sized version. ... See the **Berkeley** Enron **Email** Analysis page for more. ...
jheer.org/enron/v1/ - 25k - <u>Cached</u> - <u>Similar pages</u>

Professional Opportunities
... of **California** at **Berkeley** when he saw the e-**mail** message from India, ...
The IIT Kanpur paper, however, had run afoul of his junk **mail filter,** ...
www.siam.org/siamnews/09-02/primality.htm - 19k - <u>Cached</u> - <u>Similar pages</u>

Collaborative **Filtering** Research Papers
The collaborative **filtering algorithms** STI Pearson, STIN2, ... University of
California, Berkeley): Eigentaste is a collaborative **filtering algorithm** that ...
jamesthornton.com/cf/ - 88k - Jan 30, 2006 - <u>Cached</u> - <u>Similar pages</u>

Beyond **Algorithms**: An HCI perspective on Recommender Systems
Study participants were mostly students at the University of **California, Berkeley**.
... "Agent-Based Information **Filtering** and Recommender Systems. ...
www.rashmisinha.com/articles/BeyondAlgorithms.html - 115k - <u>Cached</u> - <u>Similar pages</u>

The Impact of Emerging Technologies: E-**Mail** You Can't Ignore ...
Going beyond existing measures, such as spam **filters** and blacklists, ... at the
University of **California, Berkeley**, who studies social exchanges online, ...
www.technologyreview.com/ InfoTech/wtr_15990,300,p1.html?PM=GO - <u>Similar pages</u>

Papers and Project Descriptions, Information **Filtering** Resources

Systems include a **mail filtering** agent, a USENET news **filter**, and a learning **...**
University of **California** at **Berkeley**: An extensive set of links to work on **...**
www.ee.umd.edu/medlab/**filter**/papers.html - 18k - Cached - Similar pages

Goooooooooogle ▶

Result Page: **1** 2 3 4 5 6 7 8 9 10   **Next**

| california berkeley mail filter algorithi | Search |

Search within results | Language Tools | Search Tips | Dissatisfied? Help us improve

Google Home - Advertising Programs - Business Solutions - About Google

©2006 Google

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	2	@ad<"20010614" and 709/206,207.ccls. and "706".clas. and ((email$2 or electornic adj3 mail$2) same filter$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 17:52
S2	35	@ad<"20010614" and (709/206,207.ccls. or "706".clas.) and ((email$2 or electornic adj3 mail$2) same filter$3) and compar$3	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 14:22
S3	62	@ad<"20010614" and (709/206,207.ccls. or "706".clas.) and ((email$2 or electornic adj3 mail$2) same filter$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 11:47
S4	10	@ad<"20010614" and (709/206,207.ccls. or "706".clas.) and (((email$2 or electornic) adj3 mail$2) same filter$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 11:47
S5	544	@ad<"20010614" and (709/206,207.ccls. or "706".clas.) and (email$2 or (electornic adj3 mail$2) same filter$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 11:51
S6	50	@ad<"20010614" and (709/206,207.ccls. or "706".clas.) and (email$2 or (electornic adj3 mail$2) same filter$3) and ((multiple or plura$4 or two) same filter$2)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 12:06
S7	7725	@ad<"20010614" and microsoft.asn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 12:06
S8	9	@ad<"20010614" and microsoft.asn. and ((email or "electronic mail") same filter)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S9	2	"20020083095"	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 12:26
S10	2	@ad<"20010614" and ((email or "electronic mail") same filter) and ((parallel or concurrent)near10 filter)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S11	2	"6578025".pn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 13:56
S12	3	"6701305".pn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 14:16
S13	126	@ad<"20010614" and (vector same compare same angle)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S14	37	@ad<"20010614" and (vector same compare same angle) and filter	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S15	0	@ad<"20010614" and (vector same compare same angle) and (filter same (email or "electronic mail"))	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 14:23

S16	5	@ad<"20010614" and (vector same compare same angle) and (filter same ((email or "electronic mail") or message))	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/01 14:23
S17	4	((email or "electronic mail") same filter) and ((parallel or concurrent)near10 filter)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S18	995	((email$2 or electornic adj3 mail$2) same filter$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S19	729	((email or "electronic mail") same filter)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S20	174	(vector same compare same angle)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S21	56	(vector same compare same angle) and filter	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2004/12/09 10:01
S32	2	"6208971".pn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 10:53
S33	278	bellegarda	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 10:53
S34	6	bellegarda.asn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:38
S35	374	LSA and message$2	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:38
S36	71	LSA and ("electronic mail" or email or e-mail$2) and message$2	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:39
S37	71	LSA and ("electronic mail" or email$2 or e-mail$2) and message$2	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:39
S38	59	LSA and ("electronic mail" or email$2 or e-mail$2) and message$2 and (valid or solicit$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:40
S39	1	LSA and ("electronic mail" or email$2 or e-mail$2) and message$2 and (valid or solicit$3) and anchor	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:40
S40	59	LSA and ("electronic mail" or email$2 or e-mail$2) and message$2 and (valid or solicit$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:42

S41	10	(latent near5 semantic) and ("electronic mail" or email$2 or e-mail$2) and message$2 and (valid or solicit$3)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:53
S42	2	"6654787".pn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/04/25 11:53
S43	4	"550886".pn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 17:53
S44	2	"5754938".pn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:04
S45	0	LSI and bellcore.asn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:04
S46	2	bellcore.asn.	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:05
S47	198	"latent semantic indexing"	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:05
S48	6	"latent semantic indexing" and (filter near10 mail)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:34
S49	0	"latent semantic" near10 vector and (filter near10 mail)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:34
S50	0	"latent semantic" same vector and (filter near10 mail)	US-PGPUB; USPAT; EPO; DERWENT	OR	OFF	2005/06/10 18:34
S51	2	"5,321,833".pn.	US-PGPUB; USPAT; EPO; DERWENT; IBM_TDB	OR	OFF	2005/06/10 18:38
S52	2	"5,301,109".pn.	US-PGPUB; USPAT; EPO; DERWENT; IBM_TDB	OR	OFF	2005/06/10 18:38